working paper
department
of economics

SPECIFICATION TESTS FOR THE MULTINOMIAL LOGIT MODEL

Jerry A. Hausman
Daniel McFadden

Number 292 October 1981

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

Comments welcome

SPECIFICATION TESTS FOR THE MULTINOMIAL LOGIT MODEL

Jerry A. Hausman
Daniel McFadden

Number 292 October 1981

Jeff Dubin, Whitney Newey, and John Rust provided research assistance. The NSF and DOE provided research support.

October 1981

Comments Welcome

SPECIFICATION TESTS FOR THE MULTINOMIAL LOGIT MODEL

by

JERRY HAUSMAN

AND

DANIEL MCFADDEN

Jeff Dubin, Whitney Newey, and John Rust provided research assistance. The NSF and DOE provided research support.

SPECIFICATION TESTS FOR THE MULTINOMIAL LOGIT MODEL

by

JERRY HAUSMAN AND DANIEL MCFADDEN

Discrete choice models are now used in a wide variety of situations in applied econometrics.[1] By far the model specification which is used most often is the multinomial logit model, McFadden (1974). The multinomial logit model provides a convenient closed form for the underlying choice probabilities without any requirement of multivariate integration. Therefore, choice situations characterized by many alternatives can be treated in a computationally convenient manner. Furthermore, the likelihood function for the multinomial logit specification is globally concave which also eases the computational burden. The ease of computation and the existence of a number of computer programs has led to the many applications of the logit model. Yet it is widely known that a potentially important drawback of the multinomial logit model is the independence from irrelevant alternatives property. This property states that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set.[2] Debreu (1960) was among the first economists to discuss the implausibility of the independence from irrelevant alternatives assumption. Basically, no provision is made for

[1]McFadden (1981) provides references to many of their uses.

[2]A "universal" logit model avoids the independence from irrelevant alternatives property while maintaining the multinomial logit form by making each ratio of probabiities a function of attributes of all alternatives, McFadden (1981). It is difficult however to give an economic interpretation of this model other than as a flexible approximation to a general functional form.

different degrees of substitutability or complimentarity among the choices. While most analysts recognize the implications of the independence of irrelevant alternatives property, it has remained basically a maintained assumption in applications.

The multinomial probit model does provide an alternative specification for discrete choice models without any need for the independence of irrelevant alternatives assumption, Hausman and Wise (1978). Furthermore, a test of the 'covariance' probit specification versus the 'independent' probit specification which is very similar to the logit specification does provide a test for the independence from irrelevant alternatives assumption. But use of the multinomial probit model has been limited due to the requirement that multivariate normal integrals must be evaluated to estimate the unknown parameters. Thus, the multinomial probit model does not provide a convenient specification test for the multinomial logit model because of its complexity.

In this paper we provide two sets of computationally convenient specification tests for the multinomial logit model. The first test is an application of the Hausman (1978) specification test procedure. The basic idea for the test here is to test the reverse implication of the independence from irrelevant alternatives property. The usual implication is to note that

if two choices exist, say car and bus in a transportation choice application, that addition of a third choice, subway, will not change the ratio of probabilities of the initial two choices. Our test here is based on eliminating one or more alternatives from the choice set to see if underlying choice behavior from the restricted choice set obeys the independence from irrelevant alternatives property. We estimate the unknown parameters from both the unrestricted and restricted choice sets. If the parameter estimates are here approximately the same, then we do not reject the multinomial logit specification. The test statistic is easy to compute since it only requires computation of a quadratic form which involves the difference of the parameter estimates and the differences of the estimated covariance matrices. Thus, existing logit computer programs provide all the necessary input to the test.

The second set of specification tests that we propose are based on more classical test procedures. We consider a generalization of the multinomial logit model which is called the nested logit model, McFadden (1981). Since the multinomial logit model is a special case of the more general model when a given parameter equals one, classical test procedures such as the Wald, likelihood ratio, and Lagrange multiplier tests can be used. Of course, we have added the requirement of the specification of an alternative model to test the original model specification. Maximum likelihood estimation of the nested logit model is considerably more difficult than for the multinomial logit model. Thus we base our test procedures on one-step asymptotically equivalent estimators to maximum likelihood. Still, new programming is

required since existing multinomial logit programs do not provide asymptotic covariance matrix estimates for the one-step estimators.

We then proceed to compare the two sets of specification test procedures for an example. We find rather unexpected results. First despite a sample size of 1000, the asymptotically equivalent classical tests differ markedly in their operating characteristics. The exact size of the Wald test is 15-25% larger than the nominal size while the exact size of the Lagrange multiplier test is approximately 15-30% smaller than the nominal size. The exact size of the likelihood ratio test is quite close to its nominal size. Furthermore, the power of the Wald test is significantly greater than the other two classical tests. This result holds also when the tests are corrected for size. Thus, the Lagrange multiplier test which is based in this application on inconsistent estimates is distinctly inferior to the Wald test which is based on asymptotically efficient estimates. Perhaps, more surprising, we find the power of one Hausman test to be comparable to that of the Wald test, even though in our example this Wald test is based on the correct alternative model. The exact size of this Hausman test is within .1% of the nominal size. When it is compared to the size corrected Wald test, it proves to be superior. Thus the often quoted asymptotic power results for local departures from the null hypothesis do not provide a reliable guide to the exact performance of our specification tests in the example despite the relatively large sample size and small departures from the multinomial logit model.

The plan of the paper is as follows. In the next section we derive the Hausman-type specification test for the multinomial logit model. The

distribution theory as well as computational considerations are discussed. In Section 3 we apply the specification test to an actual choice situation. In the next section we derive the classical tests from the nested logit model. We apply the tests to the same data as we did for the Hausman test. All four tests lead to a decisive rejection of the multinomial logit model in this application. In Section 4 we calculate the exact size and power for the two sets of specification tests for an example. Lastly, in the conclusion we report some additional empirical results and discuss some further considerations for the test procedures.

I. A Hausman-Type Test of the IIA Property

A widely used functional form for discrete choice probabilities is the multinomial logit (MNL) model

$$P(i|z, C, \beta) = e^{z_i\beta} \Big/ \sum_{j\varepsilon C} e^{z_j\beta} \tag{1.1}$$

where

$C = \{1,\ldots,J\}$ is a finite choice set

i,j = alternatives in C

z_j = a K-vector of explanatory variables describing the attributes of alternatives j and/or the characteristics of the decision-maker which affect the desirability of alternative j.

$z = (z_1,\ldots,z_J)$ the attributes of C

β = a K-vector of taste parameters

$P(i|z, C, \beta)$ = the probability that a randomly selected decision-maker when faced with choice set C with attributes z will choose i.

The MNL model has a necessary and sufficient characterization, termed independence from irrelevant alternatives (IIA), that the ratio of the probabilities of choosing any two alternatives is independent of the attributes or the availability of a third alternative, or

$$P(i|z, C, \beta) \equiv P(i|z, A, \beta)\, P(A|z, C, \beta) \tag{1.2}$$

where $i \varepsilon A \subseteq C$ and

$$P(A|z, C, \beta) = \sum_{j\varepsilon A} P(j|z, C, \beta) \tag{1.3}$$

This property greatly facilitates estimation and forecasting because it implies the model can be estimated from data on binomial choices, or by restricting attention to choice within a limited subset of the full choice set. On the other hand, this property severely restricts the flexibility of the functional form, forcing equal cross-elasticities of the probabilities of choosing various alternatives with respect to an attribute of one alternative. Further discussion of the IIA property and conditions under which it is likely to be true or false is given in Domencich and McFadden (1975) McFadden, Tye, and Train (1976), and Hausman-Wise (1978). The McFadden, Tye and Train paper suggests that the MNL specification be tested by comparing parameter estimates obtained from choice data from the full choice set with estimates obtained from conditional choice data from a restricted choice set. Here we develop an asymptotic test statistic for this comparison, using the approach to specification tests introduced by Hausman (1978).

Consider a random sample with observations $n=1,\ldots,N$. Let z^n be the attributes of C for case n, and define $S_{in} = 1$ if case n chooses i and $S_{in} = 0$ otherwise. The log likelihood of the sample is

$$L_C(\beta) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i\in C} S_{in} \ln P(i|z^n, C, \beta) \tag{1.4}$$

We first review the asymptotic properties of maximum likelihood estimates of β from (4). We make the following regularity assumptions:

a. The vector of attributes z has a distribution μ in the population which has a bounded support.
b. The MNL specification (1) with a parameter vector β^* is the true model.
c. The parameter vector β^* is asymptotically identified, i.e., if $\beta \neq \beta^*$, there exists a set Z of z values and an alternative i such that

$\int_Z P(i|z, C, \beta^*)\, d\mu(z) \neq \int_Z P(i|z, C, \beta)\, d\mu(z)$. Under these assumptions, $E\, S_{in} = P(i|z^n, C, \beta^*)$, the log likelihood converges uniformly in β to

$$\operatorname*{plim}_{N\to\infty} L_C(\beta) = \int \sum_{i\in C} P(i|z, C, \beta^*) \ln P(i|z, C, \beta)\, d\mu(z), \qquad (1.5)$$

and (5) has a unique maximum at $\beta = \beta^*$. Then the maximum likelihood estimator β_C is consistent, and $\sqrt{N}\,(\beta_C - \beta^*)$ converges in distribution to a normal random vector with zero mean and covariance matrix $\operatorname*{plim}_{N\to\infty} (-\partial^2 L_C(\beta^*) / \partial\beta\partial\beta')$. Discussion and proofs of these properties can be found in Manski and McFadden (1981); see also McFadden (1973).

Let $A = \{1,\ldots,M\}$ be a subset of the choice set C. Consider the conditional log likelihood of the subsample who make choices from A. If the MNL specification (1) is true, then the IIA property states that the probability of choosing i from C, given that the choice is contained in A, equals the probability of choosing i from A. The conditional log likelihood is then

$$L_A(\beta) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i\in A} S_{in} \ln \left(P(i|z^n, C, \beta)/P(A|z^n, C, \beta)\right) \qquad (1.6)$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i\in A} S_{in} \ln P(i|z^n, A, \beta).$$

Some components of β^*, such as the coefficients of alternative-specific

variables for excluded alternatives, are not identified by choice from A. Let $z^n = (y^n, x^n)$ be a partition of the explanatory variables into a vector y^n which only varies outside A and a vector x^n which varies within A, and let $\beta = (\gamma, \theta)$ be a commensurate partition of the parameter vector. The conditional choice probability is then

$$P(i|x^n, A, \theta) = e^{x_i^n \theta} \Big/ \sum_{j \in A} e^{x_j^n \theta} \tag{1.7}$$

and $y_i^n = y_j^n$ for i, j ε A. We add to the regularity assumptions the asymptotic identification condition

d. If $\theta \neq \theta^*$, there exists a set Z of z values and an alternative $i \varepsilon A$ such that

$$\int_Z P(i|x, A, \theta^*)\, d\mu(z) \neq \int_Z P(i|x, A, \theta)\, d\mu(z).$$

Then, as in the unconditional case, the conditional log likelihood converges uniformly in θ to

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} L_A(\theta) &= \int \sum_{i \varepsilon A} P(i|z, C, \beta^*) \ln P(i|x, A, \beta)\, d\mu(z) \\ &= \int P(A|z, C, \beta^*) \sum_{i \varepsilon A} P(i|x, A, \theta^*) \ln P(i|x, A, \theta)\, d\mu(z) \end{aligned} \tag{1.8}$$

with a unique maximum at $\theta = \theta^*$. The maximum likelihood estimator θ_A is consistent and $\sqrt{N}(\theta_A - \theta^*)$ is asymptotically normal with mean zero and covariance matrix $\operatorname*{plim}_{N\to\infty} (-\partial^2 L_A(\theta^*)/\partial\theta\partial\theta')$.

The specification test statistic is based on the parameter difference $\delta = \theta_A - \theta_C$, where $\beta_C = (\gamma_C, \theta_C)$. When the regularity assumptions hold and the MNL model is true, $\operatorname*{plim}_{N\to\infty} \delta = 0$. Conversely, when the MNL specification

(1) is false, then the IIA property fails, and (7) becomes

$$\text{plim } L_A(\theta) = \int P(A|z, C, \beta^*) \sum_{i \varepsilon A} \left[\frac{P(i|z, C, \beta^*)}{P(A|z, C, \beta^*)} \right] \ln P(i|x, A\ \theta)\ d\mu\ (z)$$

with $P(i|z, C, \beta^*)\ /\ P(A|z, C, \beta^*) \neq P(i|x, A, \theta^*)$. In general, equation (1.8) is not maximized at $\theta = \theta^*$, implying $\text{plim}_{N\to\infty}\ \delta \neq 0$. Thus, a test of $\delta = 0$ is a test of the MNL specification. Rejection of $\delta = 0$ indicates a failure of the restrictive structure of the MNL form embodied in the IIA property, or a misspecification of the explanatory variables z in (1), or both.

Acceptance of $\delta = 0$ implies that for the <u>given</u> specification of explanatory variables and distribution of these variables, the IIA property holds. Thus the test is consistent against this family of alternatives. However it is not necessarily consistent against all members of the family of alternatives defined by a given specification of explanatory variables and <u>any</u> distribution of these variables.

To derive an asymptotic test statistic for $\delta = 0$, note that under the regularity assumptions, $\sqrt{N}\ (\beta_C - \beta^*,\ \theta_A - \theta^*)$ is asymptotically normal with mean zero and a covariance matrix V calculated below; the argument is a standard application of a central limit theorem, as used in Manski and McFadden, (1981).

We require the gradients and hessians of the likelihood functions and their moments. First, the moments of S_{in} are

$$E(S_{in}) = P(i|z^n, C, \beta^*) \tag{1.10}$$

$$\mathrm{cov}(S_{in}, S_{jn}) = \begin{Bmatrix} P(i|z^n, C, \beta^*)(1-P(i|z^n, C, \beta^*)) & \text{if } i=j \\ -P(i|z^n, C, \beta^*)\, P(j|z^n, C\ \beta^*) & \text{if } i \neq j \end{Bmatrix} \tag{1.11}$$

The gradients are

$$\partial L_C(\beta)/\partial\beta = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon C} S_{in} \quad \partial \ln P(i|z^n, C, \beta)/\partial\beta$$

$$\frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon C} (S_{in} - P(i|z^n, C, \beta))(z_i^n - z_C^n) \tag{1.12}$$

where

$$z_C^n = \sum_{i \varepsilon C} z_i^n P(i|z^n, C, \beta) \tag{1.13}$$

and

$$\partial L_A(\theta)/\partial\theta = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon A} S_{in} \quad \partial \ln P(i|x^n, A, \theta)/\partial\theta$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon A} (S_{in} - S_{An} P(i|x^n, A, \theta))\ (x_i^n - x_A^n) \tag{1.14}$$

where

$$\begin{aligned} x_A^n &= \sum_{i \varepsilon A} x_i^n P(i|x^n, A, \theta), \\ S_{An} &= \sum_{i \varepsilon A} S_{in} \end{aligned} \tag{1.15}$$

The hessian for the unconditional log likelihood is

$$\begin{aligned} H_C &\equiv - \partial^2 L_C(\beta)/\partial\beta\, \partial\beta' \equiv - E\ \partial^2 L_C(\beta)/\partial\beta\partial\beta' \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon C} P(i|z^n, C, \beta)(z_i^n - z_A^n)(z_i^n - z_A^n)'. \end{aligned} \tag{1.16}$$

For the conditional log likelihood,

$$-\partial^2 L_A(\theta)/\partial\theta\partial\theta' = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon C} S_{An} P(i|x^n, A, \theta)(x_i^n - x_C^n)(x_i^n - x_C^n)' \tag{1.17}$$

and hence

$$H_A \equiv - E\, \partial^2 L_A(\theta)/\partial\theta\partial\theta'$$
$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon A} P(A|z^n, C, \beta^*)P(i|x^n, A, \theta)(x_i^n - x_A^n)(x_i^n - x_A^n)' \quad (1.18)$$

Evaluated at $\theta = \theta^*$, H_A is then

$$H_A^* = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon A} P(i|z^n, C, \beta^*)(x_i^n - x_A^n)(x_i^n - x_A^n)' \quad (1.19)$$

We now turn to calculation of the asymptotic covariance matrix V. Taylor's expansions of the gradients imply

$$\begin{bmatrix} H_C^* & 0 \\ 0 & H_A^* \end{bmatrix} \begin{bmatrix} \sqrt{N}(\beta_C - \beta^*) \\ \sqrt{N}(\theta_A - \theta^*) \end{bmatrix} = \begin{bmatrix} \sqrt{N}\ \partial L_C(\beta^*)/\partial\beta \\ \sqrt{N}\ \partial L_A(\theta^*)/\partial\beta \end{bmatrix} + \Delta \quad (1.20)$$

where Δ is a vector satisfying plim $\Delta = 0$. Using the expectations of the outer products of the gradients of the two likelihood functions we calculate $\sqrt{N}(\partial L_C(\beta^*)/\partial\beta', \partial L_A(\theta^*)/\partial\theta') \equiv g'$ is asymptotically normal with mean zero and covariance matrix

$$\lim_{N\to\infty} E\, g\, g' = \lim_{N\to\infty} \begin{bmatrix} H_C^* & \begin{vmatrix} 0' \\ H_A^* \end{vmatrix} \\ [0 \;\; H_A^*] & H_A^* \end{bmatrix} \quad (1.21)$$

where H_C^* is the H_C from equation (1.16) evaluated at $\beta = \beta^*$ and likewise H_A^* is H_A evaluated at $\theta = \theta^*$.

From (20), the asymptotic covariance matrix is

$$V = \lim_{N\to\infty} \begin{bmatrix} H_C^* & 0 \\ 0 & H_A^* \end{bmatrix}^{-1} (\lim_{N\to\infty} E\, g\, g') \lim_{N\to\infty} \begin{bmatrix} H_C^* & 0 \\ 0 & H_A^* \end{bmatrix}^{-1}$$

$$= \lim_{N\to\infty} \left[\begin{array}{cc} H_C^{-1} & H_C^{-1} \left| \begin{array}{c} 0 \\ \hline I_A \end{array} \right| \\ [0 \;\; I_A]H_C^{-1} & H_A^{-1} \end{array} \right] \tag{1.22}$$

Define $V_A = \lim_{N\to\infty} H_A^{-1}$ and

$$V_C = \lim_{N\to\infty} H_C^{-1} = \begin{bmatrix} V_{C\gamma\gamma} & V_{C\gamma\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} \end{bmatrix}, \tag{1.23}$$

where the partition is commensurate with (γ,θ). Then

$$V = \begin{bmatrix} V_{C\gamma\gamma} & V_{C\gamma\theta} & V_{C\gamma\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} & V_{C\theta\theta} \\ V_{C\theta\gamma} & V_{C\theta\theta} & V_A \end{bmatrix} \tag{1.24}$$

The asymptotic covariance matrix of $N\,(\theta_A - \theta_C)$ is then

$$Q = V_A - V_{C\theta\theta}, \tag{1.25}$$

the difference of the asymptotic covariance matrices of θ_A and θ_C. Thus the test statistic

$$T = N(\theta_A - \theta_C)' \;\; Q^t(\theta_A - \theta_C), \tag{1.26}$$

where Q^t is a generalized inverse of Q, is asymptotically distributed chi-square with degrees of freedom equal to the rank of Q. This statistic then coincides with the general specification test statistic developed by Hausman (1978) and generalized to use of singular covariance matrices by Hausman-Taylor (1981) when an efficient estimator is available under the null hypothesis.

The estimated covariance matrices for the maximum likelihood estimator θ_C and θ_A satisfy

$$\mathrm{cov}\ (\theta_C) = \left[\sum_{n=1}^{N} \sum_{i \varepsilon C} P(i|z^n,\ C, \beta_C)(z_i^n - z_C^n)(z_i^n - z_C^n) \right]_{\theta\theta}^{-1} \tag{1.27}$$

$$\underset{N \to \infty}{\mathrm{plim}}\ N\ \mathrm{cov}\ (\theta_C) = V_{C\theta\theta} \tag{1.28}$$

$$\mathrm{cov}\ (\theta_A) = \left[\sum_{n=1}^{N} S_{An} \sum_{i \varepsilon A} P(i|x^n,\ A,\ \theta_A)(x_i^n - x_A^n)(x_i^n - x_A^n)' \right]^{-1} \tag{1.29}$$

$$\underset{N \to \infty}{\mathrm{plim}}\ N\ \mathrm{cov}\ (\theta_A) = V_A \tag{1.30}$$

Therefore, an asymptotically equivalent computational formula for T is

$$T = (\theta_A - \theta_C)' \left[\mathrm{cov}\ (\theta_A) - \mathrm{cov}\ (\theta_C)\right]^t (\theta_A - \theta_C). \tag{1.31}$$

which should be quite easy to calculate using existing logit programs.

The appropriate degrees of freedom can be computed using

$$df = \mathrm{tr}\ \left[\mathrm{cov}\ (\theta_A) - \mathrm{cov}\ (\theta_C)\right]^t \left[\mathrm{cov}\ (\theta_A) - \mathrm{cov}\ (\theta_C)\right]. \tag{1.32}$$

The regularity assumptions do not exclude the possibility that Q is less than full rank; however, deficiencies will occur only for exceptional configurations of the x_i^n variables. In particular, a sufficient condition for Q to be non-singular is that

$$\begin{aligned} H^*_{C\theta\theta} - H^*_A = &\ \frac{1}{N} \sum_{n=1}^{N} \sum_{i \varepsilon D} P(i|z^n, C, \beta^*)(x_i^n - x_D^n)(x_i^n - x_D^n)' \\ &+ \frac{1}{N} \sum_{n=1}^{N} P(A|z^n, C, \beta^*)\ P(D|z^n, C, \beta^*)(x_A^n - x_C^n)(x_A^n - x_C^n)' \end{aligned} \tag{1.33}$$

have a non-singular limit, where D = C|A and

$$x_D^n = \sum_{i \varepsilon D} P(i|z^n,\ D,\ \beta^*)\ x_i^n. \tag{1.34}$$

This is generally the case if the x variables either vary within D, or

take on values within D different from their average within A.

The estimated matrix $[\text{cov}\,(\theta_A) - \text{cov}\,(\theta_C)]$ may fail to be definite in finite samples even when Q is non-singular. This does not impede calculation of the statistic (34) or carrying out the asymptotic test. However, one can form an asymptotically equivalent estimate of cov (θ_A) such that $[\text{cov}\,(\theta_A) - \text{cov}\,(\theta_C)]$ is always positive semidefinite by evaluating $P(i|x^n, A, \theta)$ in (32) at θ_C and replacing S_{An} by $P(A|z^n, C, \beta_C)$. We have occasionally found the test statistic of equation (1.32) to be negative due to lack of positive semidefiniteness in finite sample applications. Replacement by the alternative covariance matrix always leads to a small positive number. However, in no case have we found this alternative statistic to be so large as to come close to any reasonable critical value for a χ^2 test.

II. An Application

As an application of the test, we consider consumer choice of clothes dryers. The alternatives are an electric dryer, a gas dryer, or no dryer. We consider a three-alternative MNL model of this choice. A plausible alternative is that the two types of dryers are viewed as having many common unobserved characteristics in the decision whether to have a dryer. Hence, we apply the specification test by excluding the no-dryer alternative. The sample is 1408 households from the 1975 WCMS survey of residential appliance holdings and energy use patterns; detailed discussions of the sample can be found in Newman and Day (1975) and Dubin and McFadden (1980).

The variables used in the model are: (1) clothes dryer operating cost in 1975, (2) clothes dryer capital cost, (3) homeowner dummy variable, (4) number of persons in the household, (5) gas availability index. The operating costs were calculated on the assumption that each person in the household dries two loads of clothes per week. The capital costs we calculated as the mean price plus a connect charge. The connect charge was not included for the gas dryer if other gas appliances were already owned. Similarly, a connect charge for the electric dryer was not included for the electric dryer if 220 volts using appliances were already owned. It should be noted that the choice variable reflects holdings rather than purchases. Our analysis of holdings as a function of 1975 costs requires the strong assumption that these reflect expected cost at date of purchase. We assume in addition that the choice is made by the resident as an independent

decision, and not by a landlord or as part of an overall dwelling purchase decision. Finally, while the GASAV variable and capital costs reflect the probabilities that all alternatives are in fact available to the decision-makers, we are unable to identify those individuals in the sample who are captive to one of the alternatives and unable to exercise choice. These limitations of the data and the specification of the explanatory variables make it important to emphasize that our test is a joint test of the IIA structure of the MNL model and of the variable specifications, and a rejection may be due to either or both.

In the first column of Table 2.1 we present estimates of the MNL model for the three choice (unrestricted) situation. We have included choice specific dummy variales for the electric dryer and no dryer options with a normalization that the gas dryer dummy variable equals zero. Therefore for instance, homeowners are less likely to buy electric dryers than gas dryers and are even less likely still to have no dryer. Also the availability of gas raises the probability quite strongly that a gas dryer will be owned. Both operating cost and capital cost coefficients have the expected negative sign, and they are estimated quite precisely. Their estimated ratio is .81 which leads to a calculated discount rate of about 120% per year assuming a lifetime for dryers of 8 years. This discount rate seems decidedly too high.

We now re-estimate the model after deletion of the no dryer alternative. Thus we compare the choice between an electric dryer and gas dryer among the subset of households that have a dryer. If the IIA property holds true then the coefficient estimates should remain approximately the same. Under the alternative hypothesis, we might expect the coefficients to change because

gas dryers and electric dryers are closer substitutes than the no dryer choice. Estimates from the restricted choice set are reported in the second column of Table 2. Note that while the capital cost coefficient remains the same that the operating cost coefficient has almost tripled. The estimated discount rate is now about 40% which is high but in line with other estimates for appliance purchase decisions, e.g., Dubin and McFadden (1981) and Hausman (1979). The other coefficients which change markedly are the effects of the number of persons in the household. To calculate the specification test we use equation (1.31) where the dimension of θ is 6 since the 4 no dryer specific variables are eliminated. We find T = 20.9. Since under the null hypothesis T is distributed as central χ^2 with 6 degrees of freedom, we reject the MNL specification at beyond the 99 percent critical level.

In the last column of Table 2.1 we present an alternative Hausman-type specification test where we have eliminated the electric dryer option to form the restricted choice set. Since the unrestricted logit estimates are calculated from the weighted average of the binary choice set estimates we might expect the estimates from the gas dryer-no dryer choice set to also provide the basis for a test. Examination of the results in Table 2.1 indicates that the estimates differ greatly from the unrestricted estimates. In fact, the coefficient of operating cost now has the incorrect sign. We find T = 94.1. Again under the null hypothesis T is distributed as χ^2 with 6 degrees of freedom. Therefore, we reject the MNL specification at beyond the 99 percent critical level. Note, however, that the two specification tests of this section are not independent as the combined size of the two tests is not $1-(.99)^2$. We return to this subject subsequently. Since both

specification tests lead to such a decisive rejection of the MNL specification for the dryer choice problem, we may safely conclude that misspecification is a serious problem with our model. We now consider an alternative model specification to the MNL, the nested logit model which allows for more flexible patterns of similarities. We also develop an alternative specification test based on classical testing procedures. In cases where the nested logit model is the correct specification, it might be expected to provide a more powerful test from the omnibus-type Hausman test which in this application is not based on a specific parametric alternative hypothesis.

TABLE 2.1

Unrestricted and Restricted MNL Choice Models

Variable	Unrestricted Estimate (A.S.E.)	Restricted Estimate (A.S.E.)	Alternative Restricted Estimate (A.S.E.)
1. Operating cost	-.013 (.005)	-.036 (.007)	.195 (.038)
2. Capital Cost	-.016 (.001)	-.016 (.001)	-.073 (.012)
3. House owner x Elec. dryer dummy	-.631 (.280)	-.695 (.298)	-- --
4. Persons x elec. Dryer dummy	.068 (.056)	.227 (.076)	-- --
5. Gas availability x Elec. dryer dummy	-1.229 (.476)	-1.696 (.507)	-- --
6. Elec. dryer dummy	2.096 (.458)	2.616 (.494)	-- --
7. House owner x No dryer dummy	-1.691 (.263)	-- --	-1.88 (.282)
8. Persons x No dryer dummy	-.376 (.049	-- --	.117 (.090)
9. Gas availability x No dryer dummy	-1.512 (.497)	-- --	-.536 (.609)
10. No dryer dummy	.093 (.558)	--	-14.4 (2.97)
Log Likelihood	-1344.9	-498.8	-407.6
Number of observations	1408	961	809

III An Alternative Nested Logit Specification and Classical Tests

The use of the Hausman-type specification test in Sections I-II requires no specific alternative model.[1] In this section we consider a specific alternative model, the nested logit model, to base test procedures on. Given a parametric alternative hypothesis, we can apply classical test procedures such as the Wald test, likelihood ratio (LR) test, and Lagrange multiplier (LM) test. It is well known that for local deviations from the null hypothesis, these tests have certain optimal large sample power properties, c.f. Silvey (1970) and Cox and Hinckley (1974). Of course, the optimum properties of these classical tests depend on three factors which may not be satisfied in a given application: (1) the alternative specification on which these tests are based is correct, (2) the sample is large enough so that the asymptotic theory provides a good approximation, (3) deviations from the null hypothesis model are of order $1/\sqrt{N}$. For our dryer choice application of the previous section, it appears that the nested logit model provides an attractive model for the alternative specification. In many other applications the choice of the alternative model may not be nearly so clearcut. The question of sufficiently large samples and local deviations we postpone to the next section.

---

[1]It is interesting to note that the previous test is not equivalent to an ANCOVA-like procedure in which β would be allowed to vary across each alternative or some subset of alternatives after a normalization. The specification test here would involve a test of the equality of the β's. But it is straightforward to check that the IIA property still holds under this specification so that the most likely failing of the MNL model would not be tested.

The nested logit model for the three choice case has the simple hierarchial nature shown in Figure 3.1. Alternatives 1 and 2 are assumed to have more common characteristics than either alternative has with alternative 3.

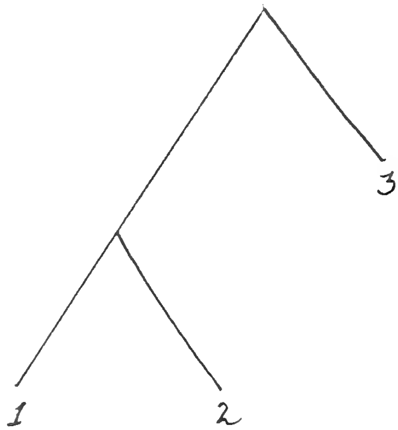

Fig. 3.1

The idea behind the choice process is that the individual forms a weighted average of the attributes of alternatives 1 and 2, sometimes called the inclusive value which is closely related to his consumer's surplus from these two choices considered above.

$$y = \log\left(e^{z_1\beta/\lambda} + e^{z_2\beta/\lambda}\right) \tag{3.1}$$

where λ is a scalar parameter of the model. The choice probabilities of the model are

$$P(1|z,C,\beta,\lambda) = p_1 = \frac{e^{z_1\beta/\lambda}\, e^{\lambda y}}{e^{y}\,\left(e^{z_3\beta} + e^{\lambda y}\right)} \tag{3.2.a}$$

$$P(2|z,C,\beta,\lambda) = p_2 = \frac{e^{Z_2\beta/\lambda}\, e^{\lambda y}}{e^{y}\,\left(e^{z_3\beta} + e^{\lambda y}\right)} \tag{3.2.b}$$

$$P(3|z,C,\beta,\lambda) = p_3 = \frac{e^{z_3\beta}}{e^{z_3\beta} + e^{\lambda y}} \tag{3.3.b}$$

For $\lambda = 1$, the nested logit model reduces to a MNL model. For $0<\lambda<1$ the model fails to satisfy the IIA property but it does satisfy the properties required for a random utility model. This proposition is proven together with a discussion of other features of the general model specification in McFadden (1981). One property of the model which deserves mention is that on any subbranch of the tree, the IIA assumption is made. However, in our application since only a binary comparison involving alternatives 1 and 2 is required, the IIA property does not lead to a testable restriction.

Given the alternative the classical test procedures may be applied by basing tests on the likelihood function

$$L_N(\beta,\lambda) = \frac{1}{N} \sum_{n=1}^{N} \sum_{\lambda\varepsilon C} S_{in} \log P(i|z^n,C,\beta,\lambda) \tag{3.4}$$

The assumptions for the MNL logit model following equation (1.4) are sufficient to prove consistency and asymptotic normality of the estimates $(\beta,\lambda) = \delta$ where we replace β by δ in the assumptions. The covariance matrix of the asymptotic normal distribution equals $\operatorname*{plim}_{N\to\infty} (-\partial^2 L_N(\delta^*)/\partial\delta\partial\delta')$.

The classical tests proceed so that each test leads to a test of $\lambda=1$, so that

under the null hypothesis the test statistic is a central χ^2 random variable with 1 degree of freedom.[1] But a potential problem arises in that the likelihood function of equation (3.4) is not inexpensive to maximize as application of both the Wald test and LR test require. A ridge in the likelihood space makes for slow progress for most algorithms. The likelihood function is not globally concave as is the MNL likelihood function. An alternative procedure would be to maximize equation (3.4) conditional on λ and to choose the maximized value of L_N with respect to both β and λ. However, each of these conditional maximizations requires an optimization routine so that the entire process is not inexpensive. Since we want to find an inexpensive test procedure for the MNL model, we instead based our tests on one Berndt-Hall-Hall-Hausman (1974) step beginning from consistent parameter estimates $(\tilde{\beta},\tilde{\lambda})$. At least since R.A. Fisher's 1925 article, it has been known that the resulting estimates are asymptotically equivalent to the ML estimates.[2]

We derive our intial consistent estimates from the so-called "sequential estimator", McFadden (1981). The sequential estimator first estimates the vector β/λ and an estimate of the inclusive value y from estimation based on the restricted choice set, A which contains alternatives 1 and 2. The estimate of $\tilde{\lambda}$ is then derived by use of a MNL program which considers the

---

[1]Here we take the alternative hypothesis to be $\lambda \neq 1$. One might consider the more restricted alternative that $\lambda < 1$ given the random utility model requirements. Since only 1 parameter is under test, the critical values for the test could be easily adjusted.

[2]Fisher considered the i.i.d. case while here the presence of the z^{n}'s lead to a i.n.i.d. case. However, the proof of asymptotic equivalence is straightforward.

choice between alternative 1 or 2 and alternative 3, along with the estimate $\tilde{\beta}$ Now in theory the BHHH step leads to asymptotically efficient estimates; in practice, we found it was a common occurence for the likelihood function to decrease relative to its value at the consistent, but inefficient, sequential estimates.

Therefore, we took two or more steps until an increase in the value of the LF occurred. The asymptotic theory remains the same although the results are somewhat worrisome in light of our sample size of nearly 1500.

The results of these procedures are given in Table 3.1. In the first column we chose the 'correct' nested model where the two dryer choices are combined on one branch of the tree. For our nested logit specification alternatives 1 and 2 are electric and gas dryers with alternative 3 being the no dryer option. The estimates of β change significantly from Table 2.1. The implied discount rate is now about 72%, in between the restricted and unrestricted logit estimates. Gas availability is now estimated to be less important, but the no dryer dummy variable has a much larger effect. In terms of our test procedures we find the following:
(1) The estimated extra parameter is $\tilde{\lambda}$ = .364 which is 5.17 asymptotic standard deviations from the MNL value of 1.0. The Wald test is therefore calculated to be 26.7 which gives a very strong rejection of the null hypothesis. (2) The LR test takes twice the difference of the log likelihood functions and equals 6.8. The MNL specification is again strongly rejected.

TABLE 3.1

Nested Logit Estimates and Classical Test Results

Variable	Gas-Electric Nested Logit Estimate (A.S.E.)	Gas-No Dryer Nested Logit Estimate (A.S.E.)
1. Operating Cost	-.0089 (.0045)	-.134 (.004)
2. Capital Cost	-.0065 (.0020)	-.146 (.007)
3. House owner x Elec. dryer dummy	-.284 (.154)	.383 (.148)
4. Persons x Elec. dryer dummy	.047 (.035)	.387 (.041)
5. Gas availability x Elec. dryer dummy	-.556 (.292)	-.730 (.271)
6. Elec. dryer dummy	.944 (.381)	-2.35 (.253)
7. House owner x No dryer dummy	-1.50 (.197)	-.763 (.061)
8. Persons x No dryer dummy	-.327 (.037)	.053 (.013)
9. Gas availability x No dryer dummy	-.919 (.441)	-.256 (.119)
10. No dryer dummy	1.01 (.397)	-5.62 (.319)
11. λ	.364 (.123)	.405 (.025)
Log Likelihood	-1341.5	-3943.8
Number of obstructions	1408	1408

(3) The LM test calculates $(\partial L/\partial\delta)'Q^{-1}(\partial L/\partial\delta)$ where Q is a consistent estimate of the covariance matrix of the asymptotic distribution of N $(\tilde{\delta}-\delta)$ under the null hypothesis. Here it is calculated to be 5.65. Thus, all three classical tests lead to a rejection of the MNL specification. They also lead to a rejection at a higher level of significance than the Hausman tests although those tests did reject at beyond the 1% level.

To further investigate the performance of the classical tests, in column 2 of Table 3.1 we present estimates of an 'incorrect' nested logit specification. Here we made the gas dryer choice correspond to choice 1 in Figure 3.1 and the no dryer option correspond to choice 2. On the other branch, the electric dryer option is choice 3. Note that the coefficient estimates change markedly, especially for those coefficients which interact with choice specific dummy variables. The calculated discount rate is now over 100 percent since the estimated coefficient of capital cost exceeds the estimated coefficient of operating cost. The results of the classical tests are: (1) The estimated parameter is $\hat{\lambda}$ = .405 which is 23.8 asymptotic standard deviations from the MNL value of 1.0. The Wald test is therefore 528.9 which is a stronger rejection of the MNL model than the 'correct' nested logit model gave. (2) The LR test turns out to be unuseable. A negative LR statistic is calculated even after three iterations at the initial "consistent" estimates. (3) The LM test is calculated to be 30.9. Thus both the Wald test and LM tests based on the 'misspecified' nested logit model lead to stronger rejection of the original MNL specification than do the corresponding tests based on the "correct" nested logit model. As with the Hausman specification tests, the two sets of tests are not independent. We next investigate whether the tests can be distinguished on the basis of

size and power characteristics. For instance, if the classical tests are found to have a large power advantage over the Hausman test, they might be the tests of choice despite their increased computational requirements.

IV. Exact and Approximate Comparisons of the Hausman Test and the Classical Tests

The results of this section may prove surprising to devotees of classical test procedures, especially recent advocates of LM tests.[1] We consider a three choice example when the Hausman test and classical tests have closed form solutions. Exact comparisons of the size and power of the tests can then be made. To make the comparisons, we choose a nested logit model as the correct model under the alternative hypothesis. Therefore, the Wald, LR, and LM tests are based on the correct alternative specification. The Hausman specification test is, of course, not based on a specific alternative model. Just these type of conditions lead to theorems of optimal asymptotic power among the Wald, LR, and LM - the so-called holy trinity - of asymptotic tests. For our particular example we choose the sample size equal to 1000 so that asymptotic theory should provide a reasonable guide. Yet we find two rather unexpected results: (1) The three tests which comprise the trinity differ markedly in both of their operating characteristics even when their nominal size is the same.[2]

---

[1]Two classes of LM tests seem worthwhile to distinguish here. One class is based on initial consistent estimates so that the LM test is equivalent to taking one Newton type step which under general regularity conditions will produce asymptotically efficient estimates. The other class of LM tests is based on initially inconsistent estimates under the alternative hypothesis, and except in the linear case or for local alternatives their finite sample properties have not proven to be generally attractive. It is the second class of LM tests that we consider in our application.

[2]That the actual size differs in a predictable way for a class of linear multivariate regression models was proven by Berndt-Savin [1978].

(2) Uncorrected for size, the Wald test and Hausman test have approximately equal power. Next comes the LR test with the LM test distinctly in the rear. Both results may well arise because we are considering a non-local alternative hypothesis.[1] That is, the expansions required for the optimal power theorems hold for parameter vectors $\theta_A = \theta_o + \delta$ where δ must be sufficiently small, e.g., $\delta = d/\sqrt{N}$ where d is a constant vector. But, examples have been given such as Peers (1971) where significant differences can arise. No generally accepted theory exists for the non-local case. Or, the samples may not be large enough for the reliable application of asymptotic theory. We see our results as a particular example and as a caution against relying too heavily on the local asymptotic theory.

For our example we consider a 3 choice MNL specification of equation (1.1) with only a single explanatory variable. Furthermore, we assume only a single data configuration occurs, $z_1=1$, $z_2=0$, $z_3=0$. We assume N=1000 repetitions of the choice and cell counts n_1, n_2, n_3. We assume that the true parameter $\beta=\log 2$ generates the observations under the null hypothesis. For these data the MNL choice probabilities take the form:

$$p_1 = e^{\beta}/(2 + e^{\beta}) \qquad p_2 = p_3 = 1/(2 + e^{\beta}) \tag{4.1}$$

which for $\beta=\log 2$ are $p_1 = .5$ and $p_2 = p_3 = .25$. The log likelihood function for the unrestricted choice set is

$$L_c(\beta) = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 = n_1 \beta - N \log (2+e^{\beta}). \tag{4.2}$$

Maximization of the likelihood function yields the estimates

[1] A discussion of local and non-local alternatives for an application of the Hausman test is given in Hausman [1978].

$$\beta_C = \log\left(\frac{2n_1}{n_2 + n_3}\right) \qquad V_C = N/n_1(n_2 + n_3) \tag{4.3}$$

where V is the large sample estimator of the variance of β given in equation (1.16).

For the restricted choice set we eliminate choice 3. Maximization of the restricted log likelihood function $L_A(\beta)$ of equation (1.6) yields estimates

$$\beta_A = \log\left(\frac{n_1}{n_2}\right) \qquad V_A = (n_1 + n_2)/n_1 n_2 \tag{4.4}$$

where V_A is calculated from equation (1.17). The Hausman test statistic based on the deletion of alternative 3 is

$$H_3 = (\beta_C - \beta_A)^2/(V_C - V_A) = \left(\log\left(\frac{2n_2}{n_2 + n_3}\right)\right)^2 n_2(n_2 + n_3)/n_3. \tag{4.5}$$

By symmetry the Hausman statistic based on the deletion of alternative 2 is

$$H_2 = \left(\log\left(\frac{2n_3}{n_2 + n_3}\right)\right)^2 n_2(n_2 + n_3)/n_3. \tag{4.6}$$

Note that the last possible test H_1 is not defined under our data configuration since β is not identified when alternative 1 is deleted.

As the model specification for the alternative hypothesis we use the nested logit model. For our uses it can be most conveniently written as

$$p_1 = \Pi_1 p \quad p_2 = \Pi_2 p \quad p_3 = 1-p = e^{\beta z_3}/(e^{\beta z_3} + (e^{\beta z_1/\lambda} + e^{\beta z_2/\lambda})^{\lambda}) \tag{4.7}$$

where $\Pi_i = e^{\beta z_i/\lambda}/(e^{\beta z_1/\lambda} + e^{\beta z_2/\lambda})$ for $i = 1,2$. The log likelihood has the form

$$L(\alpha,\lambda) = n_1\alpha - (n_1+n_2)(1-\lambda)\log(e^{\alpha}+1) - N\log(1+(e^{\alpha}+1)^{\lambda}) \tag{4.8}$$

where we use the parameterization $\alpha = \beta/\lambda$. The maximum likelihood estimates are

$$\alpha = \log \frac{n_1}{n_2} \qquad \lambda = \log \left(\frac{n_3}{n_1+n_2}\right) / \log \left(\frac{n_2}{n_1+n_2}\right). \tag{4.9}$$

Denote the partitioned large sample estimate of the information matrix as

$$-\lim_{N\to\infty} E \begin{bmatrix} L_{\alpha\alpha} & L_{\alpha\lambda} \\ L_{\alpha\lambda} & L_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\lambda} \\ A_{\lambda\alpha} & A_{\lambda\lambda} \end{bmatrix}. \tag{4.10}$$

Then the Wald statistic for the hypothesis H_0: $\lambda=1$ is

$$\hat{W} = (\lambda - 1)^2 \hat{V}^{-1} \tag{4.11}$$

where $\hat{V}^{-1} = A_{\lambda\lambda} - A_{\alpha\lambda}^2 / A_{\alpha\alpha} = \frac{n_2 n_3}{N} t_1^2 (t_1^2 + t_2^2 n_3/N)/(t_1^2 n_2/(n_1+n_2) + t_2^2 n_3/N)$

for $t_1 = \log(n_2/(n_1+n_2))$ and $t_2 = \log(n_3/(n_1+n_2))$.[1]

Next the LR statistic is calculated from the unrestricted log likelihood function of equation (4.2) for the MNL model which sets $\lambda=1$ and the nested log likelihood of equation (4.8):

$$LR = 2(L(\alpha,\lambda) - L_c(\beta)) = 2n_2 \log \left(\frac{2n_2}{n_2+n_3}\right) + 2n_3 \log \left(\frac{2n_3}{n_2+n_3}\right) \tag{4.12}$$

Finally we derive the LM statistic. Under the null hypothesis with $\lambda=1$, we

---

[1]Details of these derivations will be provided by the authors upon request.

use the MNL estimate $\beta = \log(2n_1/(n_2+n_3))$ so that in equation (4.7) $\Pi_2 = (n_2+n_3)/(N+n_1)$ and $p = (N+n_1)/2N$. We then evaluate the gradient of the likelihood function (4.8) at this point to find

$$L_\lambda = \left(\frac{n_3-n_2}{2}\right)\log (n_2+n_3)/(N+n_1) \tag{4.13}$$

together with its large sample variance

$$\begin{aligned}\tilde{\tilde{V}}^{-1} &= -\,(EL_{\lambda\lambda} - (EL_{\lambda\alpha})^2/EL_{\alpha\alpha}) \\ &= \frac{3N-n_1}{3N+n_1}\;\frac{(n_2+n_3)(N+n_1)}{4N}\;(\log((n_2+n_3)/(N+n_1)))^2.\end{aligned} \tag{4.14}$$

Therefore we calculate the LM statistic as

$$LM = L_\lambda^2/\tilde{\tilde{V}}^{-1} = N(n_3-n_2)^2(3N+n_1)/(3N-n_1)(n_2+n_3)(N+n_1) \tag{4.15}$$

Asymptotically, each of the test statistics H_3, H_2, W, LR, LM is under the null hypothesis distributed χ_1^2. Note that (n_1,n_2,n_3) has the trinomial distribution,

$$\Pr(n_1,n_2,n_3) = \frac{N!}{n_1!n_2!n_3!}\,p_1^{n_1}\,p_2^{n_2}\,p_3^{n_3}\,, \tag{4.16}$$

where p_1,p_2,p_3 are given by equation (4.1) with $\lambda=1$ under the null hypothesis, and for $0<\lambda<1$ under alternatives. For the example, we calculate numerically the exact distribution of the statistics for $\lambda=1$ and for alternative values $\lambda = (.95,.9,.85,.8,.75,.7,.6)$. These calculations permit determination of the exact sizes of each test for various nominal sizes, and of the power functions, either size-corrected or uncorrected. The means and

variances of the parameter estimates and test statistics, conditioned on positive cell counts, are also computed.

In general when closed form test statistics cannot be found, numerical calculation of exact power requires expensive Monte Carlo simulation. A "rough-and-ready" approximation to power can however be calculated as follows: Let $S(n_1,n_2,n_3)$ be the statistic in question, and let

$$\mu = S(Np_1,Np_2,Np_3), \tag{4.17}$$

where (p_1,p_2,p_3) are the choice probabilities under an alternative, be a measure of the "location" of the statistic. If S is asymptotically χ_k^2 under the null hypothesis, then under alternative S is roughly a non-central chi-square with non-centrality parameter μ, and power can be calculated from the distribution

$$P_n(X'^2_{k,\mu} \leq x) = e^{-\mu/2} \sum_{m=0}^{\infty} \frac{(-1)^m (x/2)^{\frac{k}{2}+m}}{m!(\frac{k}{2}+m)} \sum_{n=0}^{m} \frac{m!}{n!(m-n)!} \frac{(-\mu/2)^n}{\Gamma(\frac{k}{2}+n)}. \tag{4.18}$$

Table 4.1 gives the exact distribution of the five alternative test statistics under the null hypothesis: HAUS 3 is the Hausman statistic based on deletion of alternative 3, HAUS 2 the corresponding statistic for deletion of alternative 2, and WALD, LM, and LR are the Wald, Lagrange multiplier, and Likelihood ratio statistics, respectively. Table 4.2 gives the means and variances of these statistics, exact sizes of various tests, and critical levels.

Table 4.1 shows that the exact distributions of the test statistics are relatively close to their asymptotic limit. All the statistics have thicker lower tails than the chi square. The Wald test also has a thicker

upper tail, and the LM test a thinner upper tail, than the chi square where we use the upper tail to define the critical point for the tests. The conditional means and variances given in Table 4.2 also indicate that the Hausman statistics and likelihood ratio statistic have distributions close to the asymptotic limit, while the Wald statistic is distributed with a higher mean and variance and the LM statistic with a lower mean and variance. The exact sizes of the LR, HAUS 3, and HAUS 2 tests are close to their nominal sizes when the asymptotic critical level is used. However, the exact size

TABLE 4.1

Exact Cumulative Distribution Function of the Test Statistics

ARGUMENT	CHI SQUARE	HAUS 3	HAUS 2	WALD	LM	LR
.0001562	.0100000	.0178485	.0178485	.0178485	.0178485	.0178485
.0009766	.0250000	.0178485	.0178485	.0178485	.0178485	.0178485
.0039119	.0500000	.0535097	.0535097	.0535097	.0535097	.0535097
.0157204	.1000000	.0890643	.0890643	.0890642	.0898139	.0890643
.0356288	.1500000	.1595404	.1595404	.1551770	.1595720	.1595490
.0639752	.2000000	.1943906	.1943906	.1943880	.1986097	.1943891
.1012512	.2500000	.2571302	.2571302	.2333807	.2625366	.2574044
.1481302	.3000000	.2964287	.2964287	.2956866	.3088109	.2963552
.2055111	.3500000	.3533769	.3533769	.3334747	.3615748	.3547693
.2745780	.4000000	.3964012	.3964012	.3910138	.4110970	.3948666
.3568926	.4500000	.4514412	.4514412	.4352908	.4667838	.4527005
.4545310	.5000000	.5000545	.5000545	.4849164	.5174321	.5005644
.5702946	.5500000	.5501913	.5501913	.5342629	.5663164	.5490352
.7080494	.6000000	.6011664	.6011664	.5828786	.6169712	.5987006
.8732966	.6500000	.6506719	.6506719	.6327198	.6672093	.6490455
1.0741902	.7000000	.6995021	.6995021	.6832754	.7179841	.7008306
1.3235021	.7500000	.7501350	.7501350	.7341925	.7682552	.7500117
1.6428286	.8000000	.8004872	.8004782	.7847090	.8171284	.8003099
2.0730254	.8500000	.8504647	.8504645	.8360386	.8652888	.8501056
2.7067207	.9000000	.9001828	.9001823	.8881028	.9120306	.9004682
3.8431482	.9500000	.9499366	.9499344	.9419375	.9570080	.9503231
5.0258684	.9750000	.9748535	.9748486	.9696143	.9782958	.9754136
6.6369923	.9900000	.9898452	.9898323	.9870385	.9907046	.9904561
10.8291046	.9990000	.9989388	.9989182	.9984343	.9985983	.9992103
15.1371548	.9999000	.9998936	.9998833	.9998088	.9997743	.9999600
19.5106637	.9999900	.9999924	.9999901	.9999792	.9999684	.9999999
23.9262119	.9999990	1.0000000	.9999998	.9999989	.9999975	1.0000000
28.3710278	.9999999	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000

TABLE 4.2

Characteristics of the Test Statistics

	CHI SQUARE	HAUS 3	HAUS 2	WALD	LM	LR
Mean[1]	1.0	0.99990	1.00009	1.07102	0.92832	0.99554
Variance[1]	2.0	2.01212	2.01608	2.30900	1.69288	1.95064
Exact size of a nominal 10% test[2]	0.10	0.09982	0.09982	0.11190	0.08797	0.09953
Exact size of a nominal 5% test	0.05	0.05006	0.05007	0.05806	0.04299	0.04968
Exact size of a nominal 1% test[3]	0.01	0.01015	0.01017	0.01296	0.00929	0.00954
Critical level for an exact 5% test.	3.84315	3.84528	3.84535	4.11628	3.57253	3.82318

[1] Mean and variance are conditioned on the event that all cell counts exceed 50, which occurs with probability $1-10^{-48}$.

[2] Critical level 2.70672.

[3] Critical level 5.02587.

of the wald test exceeds its nominal size substantially, while the exact size of the LM test is substantially lower.[1] For instance, for a nominal 5% test size, the Hausman tests are within 1% of the nominal size and the LR test is within .6%. But the Wald test size is off by 15.8% and the LM test by 15.1% in the other direction. For a nominal 1% test the discrepancy of the Wald test is 26%. Table 4.2 gives the adjustments in critical levels required to achieve an exact 5 percent test with each statistic.

Table 4.3 gives the power of each of the tests against the nested logit model with values of λ less than one. Power curves are given for both the asymptotic tests and the size corrected 5 percent tests. Also given are the rough estimations to power based on equations (4.17) and (4.18).

The exact powers of the asymptotic tests, uncorrected for size, are ranked

$$\text{WALD} > \text{HAUS 2} > \text{LR} > \text{LM} > \text{HAUS 3}$$

over most values of λ. The performance of the WALD and LM tests is explained in part by the deviations between the exact and nominal sizes of these tests. The differences in power are of sufficient size and uniformly to suggest that the WALD and HAUS 2 tests are clearly preferable to the remainder. The LR and LM tests do not do nearly as well by comparison.

Table 4.3, III gives the powers of the size corrected tests. Here the rankings are

---

[1]These sizes results for the WALD, LR, and LM tests are in accord with the Berndt-Savin [1978] results for multivariate linear models.

TABLE 4.3

Exact Power of the Tests

I. Asymptotic test, nominal size 0.10

	Alternative λ						
	1.0	0.95	0.90	0.85	0.80	0.75	0.70
HAUS 3							
exact	.0998	.1447	.3437	.6078	.8434	.9604	.9950
approx.	.0999	.1554	.3534	.5993	.8190	.9402	.9871
HAUS 2							
exact	.0998	.1685	.3893	.6544	.8714	.9700	.9966
approx.	.0999	.1583	.3817	.6659	.8937	.9824	.9989
WALD							
exact	.1119	.1829	.4091	.6728	.8817	.9733	.9971
approx.	.0999	.1624	.3992	.6911	.9103	.9870	.9994
LM							
exact	.0879	.1416	.3445	.6094	.8445	.9608	.9950
approx.	.0999	.1531	.3507	.6056	.8349	.9541	.9929
LR							
exact	.0995	.1564	.3669	.6320	.8583	.9656	.9959
approx.	.0999	.1568	.3670	.6316	.8579	.9648	.9955

Table 4.3 (continued)

II. Asymptotic test, nominal size = 0.05

	Alternative λ						
	1.0	.95	.90	.85	.8	.75	.7
HAUS 3							
	.0501	.0778	.2278	.4730	.7475	.9210	.9869
	.0500	.0885	.2438	.4745	.7245	.8928	.9722
HAUS 2							
	.0501	.1015	.2826	.5419	.7998	.9438	.9919
	.0500	.0906	.2682	.5451	.8242	.9633	.9971
WALD							
	.0581	.1130	.3029	.5650	.8158	.9501	.9931
	.0500	.0936	.2835	.5728	.8479	.9719	.9981
LM							
	.0429	.0779	.2332	.4807	.7540	.9240	.9876
	.0500	.0869	.2415	.4810	.7449	.9148	.9837
LR							
	.0497	.0894	.2557	.5090	.7757	.9336	.9897
	.0500	.0895	.2555	.5082	.7744	.9324	.9891

Table 4.3 (continued)

III. Size Corrected Test, exact size = 0.05

	Alternative λ						
	1.0	.95	.90	.85	.80	.75	.70
HAUS 3							
	.0500	.0775	.2272	.4722	.7471	.9208	.9868
	.0499	.0884	.2436	.4723	.7243	.8927	.9722
HAUS 2							
	.0500	.1015	.2826	.5418	.7997	.9437	.9918
	.0499	.0905	.2697	.5449	.8241	.9632	.9971
WALD							
	.0500	.1014	.2823	.5413	.7992	.9435	.9918
	.0425	.0821	.2607	.5458	.8312	.9672	.9977
LM							
	.0498	.0896	.2562	.5098	.7764	.9340	.9898
	.0588	.0992	.2641	.5091	.7669	.9252	.9864
LR							
	.0502	.0901	.2570	.5107	.7770	.9342	.9899
	.0506	.0904	.2571	.5103	.7760	.9931	.9892

Table 4.3 (continued)

IV. Asymptotic test, nominal size 0.01

	Alternative λ						
	1.0	.95	.90	.85	.80	.75	.70
HAUS 3							
	.0102	.0166	.0755	.2248	.4911	.7646	.9366
	.0134	.0208	.0967	.2475	.4924	.7343	.9028
HAUS 2							
	.0102	.0321	.1283	.3251	.6111	.8479	.9669
	.0134	.0215	.1104	.3064	.6243	.8797	.9838
WALD							
	.0130	.0381	.1450	.3529	.6396	.8648	.9720
	.0134	.0225	.1193	.3316	.6600	.9022	.9888
LM							
	.0072	.0188	.0856	.2458	.5187	.7855	.9450
	.0134	.0203	.0954	.2526	.5174	.7747	.9359
LR							
	.0095	.0235	.1007	.2753	.5548	.8112	.9545
	.0134	.0212	.1032	.2748	.5552	.8100	.9533

HAUS 2 > WALD > LR > LM > HAUS 3.

In these rankings, HAUS 2 and WALD have comparable power, and are substantially better than the LR and LM tests, which have comparable power.

The high exact power of the HAUS 2 test is surprising in view of the fact that the Wald, LM and LR tests are specific against the alternative model generating the observations. This outcome may be special to this example where the Hausman tests have the same degrees of freedom as the trinity. However, the result also indicates that for this problem the sample size of 1000 is not sufficient for the asymptotic equivalency of the trinity and second order efficiency of the LR test to take effect in the sense that the largest deviation $\lambda_A = (1.7)/\sqrt{N} = .009$ is not sufficiently local for the asymptotic power rankings to apply.

The rough approximation to power is within .003 of the exact power of all these tests, and the ranking of the tests by approximte power usually coincides with the exact power ranking. We conclude that the rough power calculation is potentially a very useful tool for quick comparisons of tests or judgments on small sample power, without the great expense of Monte Carlo calculations.

For comparison purposes, we then recalculated the exact classical tests based on a misspecified nested logit model. That is, in eqution (4.7) we interchanged p_2 and p_3 so that choices 1 and 3 now lie on the same branch. However, we continue to conduct the test procedures based on the original model specification where choices 1 and 2 were on the same branch of Figure 3.1. In Table 4.4 we present both approximate and exact size corrected test

TABLE 4.4

Approximate and Exact Test Results at Test Size .05

	Wrong Model			Correct Model		
	λ = .95	λ = .9	λ = .85	λ = .95	λ = .90	λ = .85
Nominal size 0.05						
HAUS 3	.102	.283	.542	.078	.228	.473
HAUS 2	.078	.228	.473	.102	.283	.542
WALD	.089	.250	.501	.113	.303	.565
LM	.078	.233	.481	.078	.233	.481
LR	.089	.256	.509	.089	.256	.509
Exact size 0.05						
HAUS 3	.101	.283	.542	.077	.227	.472
HAUS 2	.077	.227	.472	.101	.283	.542
WALD	.078	.228	.473	.101	.282	.541
LM	.090	.256	.510	.090	.256	.510
LR	.090	.257	.511	.090	.257	.511
MLE estimate of λ	1.050	1.112	1.171			

results for the Hausman specification test and classical test procedures. The two Hausman tests interchange operating characteristics since our example is symmetric with respect to them. Therefore, HAUS 3 is now the more powerful test. The Wald test which had excellent operating characteristics when based on the correct model now loses substantial power. Note that the Wald test is now based on inconsistent estimates. The power characteristics of the LR and LM tests are symmetric with respect to the models. Thus, the HAUS 2 is now significantly more powerful than the classical tests. The ranking based on exact sizes is

$$\text{HAUS 3} > \text{LR} > \text{LM} > \text{WALD} > \text{HAUS 2}$$

The HAUS 3 has approximately 10% more power than the LR and LM tests and 15% more power from the Wald test. Thus, it seems useful to investigate different tree structures when using the classical tests because of their sensitivity to model specifications. In the last line of Table 4.4 we also present the maximum likelihood estimates of λ which all exceed the theoretical maximum of 1.0 which again indicates misspecification. Likewise, it seems useful to base the Hausman specification test on different restricted choice sets since its power also depends on the model specification.

V. Conclusions

The finding that the Hausman test has better power than the LR or LM test for our example may make further evidence on the test statistics of interest. We also engaged in very limited Monte Carlo examples on the original data for the application of Sections 2 and 3. We took the unrestricted MNL estimates of Table 2.1 to be the true values for the β's and then generated choice outcomes using the nested logit model of equation (3.2) to be the true model. For a range of λ's we simulated choice behavior via the generalized extreme value distribution, McFadden (1981),

$$F(\varepsilon_1,\varepsilon_2,\varepsilon_3) = \exp\left[-\exp\left(\left[(-\varepsilon_1)^{1/\lambda} + (-\varepsilon_2)^{1/\lambda}\right]^{\lambda} - \varepsilon_3\right)\right] \tag{5.1}$$

and then assigned the choice for each individual to the alternative with the highest utility calculated by the random utility model, $u_j=z_j\beta$ for j=1,3. We then compared the Hausman test of Section 2 with 6 degrees of freedom to the trinity of tests from Section 3 which all have 1 degree of freedom. The values of λ that we chose were λ = (.25,.5,.75,.9,.95,1.0). At the 5% level all tests rejected for λ = (.25,.5,.75). For λ = (.9,.95,1.0) none of the tests rejected. As in our example of the previous section the Wald test always achieved the highest value (after we normalized the Hausman test) when it was based on the correctly specified model. But we cannot correct for size here since we do not have exact values. Additionally, we generated samples using the multinomial probit model of Hausman-Wise (1978).

A somewhat disturbing result here is that while the Hausman test does not reject the independent probit specification, which is quite similar to MNL with IIA except for differences in the extreme tails of the underlying distributions, the trinity of test rejects resoundingly with the statistics above 5.0. We do not yet have a satisfactory explanation for this outcome.

In terms of applying the Hausman test or the trinity of tests, the non-uniqueness of application can arise. For instance in the 3 choice case any of the 3 alternatives can be dropped or 3 different tree structures can be defined. As the number of choices grows, the different possible combinations grow factorially. If more than one test is performed, the problem of controlling for the size arises because the tests will not be independent. In the case of the Hausman test the different estimates could be combined to form a test after their joint asymptotic covariance matrix is calculated which would not be difficult[1]. Alternatively, using the IIA property of the null hypothesis, the restricted choice set A could be successively decreased when more than 3 choices exist and the size controlled for as we do when independent F tests in linear models are used. However, we have no prescription about optimal procedures in any given application.

We conclude that the Hausman test and Wald test are the best choices to test the IIA assumption in MNL models. The Wald test does require maximum likelihood, or at least asymptotically efficient estimates, of the nested logit model and correct specification of the tree structure in the nested

---

[1]We attempted such a combination for our example in Section 4. But for that example we found the tests to have perfect negative correlation. Therefore, it appears that it is possible to increase the power indefinitely for any nominal size test by choosing the appropriate weighted average of the test statistics. We leave to further investigation the problem of optimal combination of the tests.

logit model. The Wald test has higher power in all our examples than the LR or LM test for the correct specification. However, for an incorrect specification the LR and LM tests are superior to the Wald test. But the Wald test has significantly greater computational requirements than does the Hausman test. An alternative test based on a consistent estimate of λ from the sequential logit estimator is possible by the use of Neyman's [1959] C_α procedure. However, our experience with the sequential logit estimator is that it gave quite unreliable estimates of λ method for testing.[1] Since the Hausman test gave results in general close to that of the Wald test across the range of our examples, we recommend it as a general purpose specification test for the MNL model. The Wald test should also be considered when the analyst feels that the nested logit model provides the correct specification for the choice problem under consideration. The Wald test requires more sophisticated computer software; furthermore, once it is size corrected it no longer is superior in our example to the Hausman test which is very accurate with respect to its size. Furthermore, for the case of a misspecified nested logit model, the alternative Hausman test performed much better than the Wald test. But certainly some test of the IIA property should be made when the MNL model is used. Our experience is that the Hausman test has rejected the MNL specification in a number of applications.

---

[1]An alternative indication of the unreliability is that in the majority of the cases, one BHHH step beginning at the sequential logit estimates led to a decrease in the value of the likelihood function. Since the C_α test is based on the one step methodology, we decided against its use. Furthermore, in the misspecified model case, we were often unable to find an increase in the likelihood function even when up to three BHHH steps were made.

REFERENCES

Berndt, E., B. Hall, R. Hall, and J. Hausman (1974), "Estimation and Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement, 4.

Berndt, E. and G. Savin (1977), "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45.

Cox, D.R. and D. Hinckley (1974), Theoretical Statistics, London: Allen and Unwin.

Debreu, G. (1960), "Review of R. Luce, Individual Choice Behavior", American Economic Review, 50.

Domencich, T. and D. McFadden (1975), Urban Travel Demand, Amsterdam: North Holland.

Dubin, J. and D. McFadden (1980), "An Econometric Analysis of Residential Electrical Appliance Holdings and Usage", working paper, Department of Economics, M.I.T.

Hausman, J. (1978), "Specification Tests in Econometrics", Econometrica 46.

Hausman, J. (1979), "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables", Bell Journal of Economics, 10.

Hausman, J. and W. Taylor (1981), "The Relationship between Classical Tests and Specification Tests", MIT mimeo.

Hausman, J. and D. Wise (1978), "A Conditional Probit Model for Qualitative Choice", Econometrica, 46.

Manski, C. and D. McFadden (1981), "Alternative Estimators and Sample Designs for Discrete Choice Analysis" in C. Manski and D. McFadden (eds) Structural Analysis of Discrete Data, Cambridge: MIT Press.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior" in P. Zarembka (ed.) Frontiers in Econometrics, New York: Academic Press.

McFadden, D., W. Tye, and K. Train (1976), "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model", Transportation Research Record 637.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice", in C. Manski and D. McFadden (eds), Structural Analysis of Discrete Data, Cambridge, MIT Press.

Newman, D., and D. Day (1975), The American Energy Consumer, Cambridge: Ballinger.

Neyman, J. (1959), "Optimal Asymptotic Tests of Composite Statistical Hypotheses", in U. Grenander, ed., Probability and Statistics, Stockholm: Almquist and Wiksell.

Peers, H. (1971), "Likelihood Rates and Associated Test Criteria", Biometrika, 58.

Silvey, S. (1970), Statistical Inference, London: Penguin.

Date Due

[redacted]	
[redacted]	FEB. 06 1983
	OCT. 18 1994
MAY 30 1986	
31 '87	MAY 30 1985
	AUG. 31 1985
MY 07 '88	[illegible]
MAY 26 '89	DEC. [illegible]
JAN 18 1989	
MAR 30 1990	
JUL 05 1990	

Lib-26-67